2012 2nd International Conference on Management and Artificial Intelligence
IPEDR Vol.35 (2012) © (2012) IACSIT Press, Singapore

Parameter Identification and Selection for Childhood Obesity Prediction Using Data Mining

Muhamad Hariz B. Muhamad Adnan[1+], Wahidah Husain[2] and Nur`Aini Abdul Rashid[3]

School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

Abstract. The accurate identification and selection of useful parameters for childhood obesity prediction are very important. This study aims to identify childhood obesity prediction parameters for children in Malaysia, and presents the methods used to identify and select the parameters from the children's attributes, lifestyle, family, and environment. The study comprises four stages: risk factor review, data collection, parameter identifications and selection, and evaluation. Base on the results, 19 parameters were identified. The accuracy of childhood obesity prediction using the proposed parameters was 21% greater compared to a set of parameters used in a previous study.

Keywords: parameters, childhood obesity prediction, data mining.

1. Introduction

Childhood obesity is defined as an excessive accumulation of body fat and has become a common issue for children nowadays [1]. In order to reduce the incidence of childhood obesity, a number of attempts at prediction have been made. One method of predicting childhood obesity is to use data mining techniques [2-4]. Data mining techniques are able to produce good prediction output but depend on correct parameters for prediction. Therefore, improving the identification and selection of parameters will result in better prediction accuracy.

Potential childhood obesity prediction parameters consist of the characteristics of the children and their lifestyle, environment and family. The parameters can be divided into three groups: children parameters, lifestyle parameters, and family/environment parameters. Children parameters are obesity risk factors and protective factors that relate to the child himself, such as characteristics, behavior, and health [5-10]. Lifestyle parameters relate to the child`s lifestyle, such as watching television, playing computer or video games, engaging in physical activities, daily nutrition, and the way the child is raised [8, 9, 11-21]. Family/environment parameters relate to the child's family and environment, such as parental obesity, parents' education, family income, number of siblings, and the living environment[5, 7, 8, 11, 22, 23].

This study is conducted to investigate and identify risk factors and protective factors of childhood obesity from the children, lifestyle, and family/ environment. The objective is to identify suitable parameters for childhood obesity prediction using data mining.

2. Methods

2.1. Risk Factor Review

The first stage of the study was conducted using secondary sources such as journal articles, papers, conference proceedings, books, websites, and online multimedia. The purpose was to gain as much understanding as possible of childhood obesity on a global scale before identifying the factors of childhood

+ [1]mhma.com08@student.usm.my, [2]wahidah@cs.usm.my, [3]nuraini@cs.usm.my

obesity in the Malaysian community. The knowledge gained from the risk factor review was used to construct a questionnaire.

2.2. Data Collection

In this stage, questionnaires were distributed to primary school children 9 to 11 years old in Malaysia. The schools that were chosen as participants are located in a suburban area and an urban area. A total of 140 questionnaires were given to randomly selected children for the children and their parents to answer. Most of the data collected consisted of historical data. The target group was primary school children that were old enough to properly answer the questionnaires with the help of their parents. The data obtained from the questionnaires were processed and cleaned, including removal of missing values and distortions, to identify suitable candidate risk factors and protective factors.

2.3. Parameter Identification and Selection

In this stage we identified risk factors and protective factors of childhood obesity based on the collected data. The factors were divided into three categories: children factors, lifestyle factors, and family/environment factors. Parameters identified from the children factors were gender, catch-up growth, adiposity rebound, and premature birth. The data collected categorized by gender is shown in Table 1. Normal status in Table 1 refers to the children that were normal, obesity refers to the children that were obese, overweight refers to the children that were overweight, and underweight refers to the children that were underweight.

The table shows that female child obesity prevalence is 16.7% while male child obesity prevalence is 34.3%. This indicates that male children have twice the risk of obesity compared to their female counterparts. Looking at the overweight cases, the overweight prevalence for female children is 6.7% while the overweight prevalence for male children is 6.3%. This indicates that a female child has a slightly higher chance of being overweight than a male child.

The risk factors review also showed that catch-up growth, a period of very rapid weight gain during which the child's birth weight doubles, is identified as a risk factor [7, 8]. Table 2 shows the children categorized by catch-up growth occurrence.The percentage of the catch-up children that are obese is 23.5%.

Table 1 Children categorized by gender.

		Gender		Total
		Female	Male	
Status	Normal	38	10	48
	Obesity	10	11	21
	Overweight	4	2	6
	Underweight	8	9	17
Total		60	32	92

Table 2 Children categorized by catch-up growth.

		Catch-up growth		Total
		No	Yes	
Status	Normal	40	8	48
	Obesity	16	4	20
	Overweight	6	0	6
	Underweight	12	5	17
Total		74	17	91

The parameters identified from the lifestyle factors are duration of breastfeeding, duration of sleep, eating junk food, eating fried food, eating fruit, eating snacks in front of the TV, duration of watching TV, eating hot meals for supper, physical activity, eating soup and sandwich bought outside the home, and eating snacks and chocolates bought outside the home.

Table 3 shows the children categorization based on eating junk food. Based on Table 3, the probability of being overweight for children who do not eat junk food regularly is 1.7% while the probability of being overweight for children who eat junk food more regular is 15.6%. This is a significant increase in the risk of being overweight for those with a junk food eating habit. From the risk factor review, unhealthy food and snacks are a factor of childhood obesity and overweight [19-21].

In addition, a child`s tendency to eat soup, sandwiches, snacks, and chocolate bought outside the home are associated with obesity and overweight. Table 4 illustrates whether the children bought and ate soup or sandwiches outside the home more than three times a week. From Table 4, the probabilities of obesity and overweight for children who did so are 37% and 16% respectively. The obesity and overweight probabilities

for those who did not are 19.4% and 4% respectively. This shows that obesity and overweight risk increases when the children buy and eat soup or sandwiches at school more than 3 times a week.

Table 3 Children categorized by regularly eating junk food

		Eat junk food		Total
		No	Yes	
Status	Normal	32	16	48
	Obesity	13	7	20
	Overweight	1	5	6
	Underweight	13	4	17
Total		59	32	91

Table 4 Children categorized by regular soup and sandwiches.

		Buy and eat soup/sandwich more than 3 times a week		Total
		No	Yes	
Status	Normal	40	7	47
	Obesity	14	7	21
	Overweight	3	3	6
	Underweight	15	2	17
Total		72	19	91

The parameters identified from family and environment factors are mother body mass index (BMI), father BMI, parents' weight and number of children. From the risk factor review, parental obesity is identified as a risk factor of childhood obesity due to inherited genes from the parent and to family lifestyle [5, 14, 20, 24]. An obese father or mother results in a 2.5 and 4.25 obesity odds ratio respectively for their children at age 7 [8]. If both parents are obese, the odds ratio is 10.44 [8]. Normal weighted parents are not associated with childhood obesity [8].

Table 5 shows the children categorization by their mother`s weight. The table shows that the probability of obesity for children with a normal weighted mother is 16%. The probability of obesity for children with an overweight mother is 35%. The probability of obesity for children with an obese mother is 43%. This shows that obese and overweight mothers are associated with childhood obesity.

Table 6 illustrates the children categorization by their father`s weight. The table shows that the probability of obesity for children with a normal weighted father is 15%. The probability of obesity for children with an overweight father is 37%. The probability of obesity for children with an obese father is 28%. This shows that obese and overweight fathers are associated with childhood obesity.

Table 5 Children categorized by mother`s weight.

		Mother weight status				Total
		Normal	Obese	Overweight	Underweight	
Status	Normal	27	3	10	2	42
	Obesity	8	3	8	0	19
	Overweight	6	0	0	0	6
	Underweight	8	1	5	3	17
Total		49	7	23	5	84

Table 6 Children categorized by father`s weight.

		Father weight status				Total
		Normal	Obese	Overweight	Underweight	
Status	Normal	23	2	14	1	40
	Obesity	6	2	11	0	19
	Overweight	1	2	1	2	6
	Underweight	11	1	4	1	17
Total		41	7	30	4	82

The identified parameters are summarized in Table 7. These parameters were used in childhood obesity prediction using Naïve Bayesian classifier. The result were compared with those using the same classifier but a different set of parameters proposed by Zhang et al. for childhood obesity prediction [3]. Their parameters are gender, birth weight, length at 6 weeks old, time of gestation, weight gain, and BMI at 8 months and 2 years old including height.

Table 7 Selected parameters for childhood obesity prediction.

Category	Parameters
Children	Catch-up growth, adiposity rebound, premature birth, and gender
Lifestyle	Duration of breastfeeding, duration of sleep, eating junk food, eating fried food, eating fruit, eating snacks in front of TV, duration of watching TV, eating warm meals for supper, physical activity, eating soup and sandwich bought outside home, and eating snacks and chocolates bought outside home.
Family/environment	Mother BMI, father BMI, parents' weight and the number of children.

3. Results

Table 8 shows the results of the predictions using this study's proposed parameters and Zhang et al.'s parameters. The predictions using the proposed parameters resulted in 21% more accurate prediction for obese cases. But for overweight cases, the proposed parameters did not result in higher accuracy and for normal cases predictions were better using Zhang et al.'s parameters.

Table 8 The results using the proposed parameters and Zhang et al [3].

Prediction class	Zhang parameters	Proposed parameters
Obese (%)	50	71
Overweight (%)	50	50
Normal (%)	70	50

4. Discussions

The new set of parameters was proposed in order to increase the accuracy of predicting obesity. Identifying overweight and obesity risk groups is very difficult because there are no certain factors that cause them. One child might be obese, another overweight, and a third normal even though they have the same factors. Therefore the selection of the parameters used for prediction is very important.

The results show that the proposed parameters have increased the detection of children in the obesity risk group compared to Zhang et al.'s parameters. This indicates that the proposed parameters are significant for childhood obesity prediction. Further studies to fully utilize the parameters for childhood obesity prediction should be done, since they are still raw parameters where the interdependencies between parameters should be indentified for better accuracy.

5. Conclusion

The prediction of childhood obesity is a difficult task. The selection of correct parameters is important for accurate prediction. Nineteen parameters were identified in this study. Using the proposed parameters, the detection of childhood obesity risk groups improved by 21%.

6. Acknowledgements

The authors would like to thank the Universiti Sains Malaysia (USM) for supporting this study.

7. References

[1] Adnan, M.H.M., W. Husain, and N.A.A. Rashid. *A Framework for Childhood Obesity Classifications and Predictions using NBtree*. in *7th International Conference on IT in Asia (CITA11)*. 2011. Kuching, Sarawak: IEEE.

[2] Shrestha, P.N., *Applying soft computing to early obesity prediction*. 2006: Southern Illinois University Carbondale.

[3] Shaoyan Zhang, C.T., Xiaojun Zeng, Hong Qiao, Iain Buchan, John Keane, *Comparing data mining methods with logistic regression in childhood obesity prediction.* Information Systems Frontiers, 2009. **11**(4): p. 51.

[4] Bojan Novak, M.B. *Application of artificial neural networks for childhood obesity predicition*. in *ANNES '95 Proceedings of the 2nd New Zealand Two-Stream International Conference on Artificial Neural Networks and*

Expert Systems 1995: IEEE Computer Society.

[5] R. V. Kries, et al., *Maternal smoking during pregnancy and childhood obesity.* American Journal of Epidemiology, 2002. **156**(10): p. 954-961.

[6] P H Casey, et al., *Evolution of obesity in a low birth weight cohort.* Journal of Perinatology, 2011.

[7] Stettler, N., et al., *Infant weight gain and childhood overweight status in a multicenter, cohort study.* Pediatrics, 2002. **109**(2): p. 194-199.

[8] Reilly, J.J., et al., *Early life risk factors for obesity in childhood: cohort study.* BMJ, 2005. **330**(7504): p. 1357.

[9] Ebbeling, C.B., D.B. Pawlak, and D.S. Ludwig, *Childhood obesity: public-health crisis, common sense cure.* The Lancet, 2002.

[10] M. F. Rolland-Cachera, et al., *Early adiposity rebound: causes and consequences for obesity in children and adults.* International Journal of Obesity, 2006. **30**.

[11] Ong, K.K., et al., *Size at Birth and Early Childhood Growth in Relation to Maternal Smoking, Parity and Infant Breast-Feeding: Longitudinal Birth Cohort Study and Analysis.* Pediatric Research, 2002. **52**(6): p. 863-867.

[12] Apfelbacher, C.J., et al., *Predictors of overweight and obesity in five to seven-year-old children in Germany: results from cross-sectional studies.* BMC Public Health, 2008. **8**: p. 171.

[13] *Can* breastfeeding *help prevent overweight children and adolescents.* Bandolier.

[14] Prevention *of pediatric overweight and obesity.* Pediatrics, 2003. **112**(2): p. 424-430.

[15] August, G.P., et al., *Prevention and treatment of pediatric obesity: an endocrine society clinical practice* guideline *based on expert opinion.* The Journal of Clinical Endocrinology & Metabolism, 2008. **93**(12): p. 4576-4599.

[16] Rey-Lopez, J.P., et al., *Sedentary behaviour and obesity development in children and adolescents.* Nutr Metab Cardiovasc Dis, 2008. **18**(3): p. 242-51.

[17] Ludwig, D.S., K.E. Peterson, and S.L. Gortmaker, *Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis.* The Lancet, 2001. **357**(9255): p. 505-508.

[18] Kuriyan, R., et al., *Television viewing and sleep are associated with overweight among urban and semi-urban South Indian children.* Nutr J, 2007. **6**: p. 25.

[19] Hiroki Sugimori, et al., *Analysis of factors that influence body mass index from ages 3 to 6 years: a study based on the toyama cohort study.* Japan Pediatric Society 2004. **46**(3): p. 18.

[20] Barlow, S.E., *Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report.* American Academy of Pediatrics, 2007. **120 Suppl 4**: p. S164-92.

[21] Danielzik, S., et al., Parental overweight, socioeconomic status and high birth weight are the major determ*inants of overweight and obesity in 5-7 y-old children: baseline data of the Kiel Obesity Prevention Study (KOPS).* Int J Obes Relat Metab Disord, 2004. **28**(11): p. 1494-502.

[22] Kleiser, C., et al., *Potential determinants of obesity among children and adolescents in Germany: results from the cross-sectional KiGGS Study.* BMC Public Health, 2009. **9**: p. 46.

[23] Brophy, S., et al., *Risk factors for childhood obesity at age 5: analysis of the millennium cohort study.* BMC Public Health, 2009. **9**: p. 467.

[24] Whitaker, R.C., et al., *Predicting obesity in young adulthood from childhood and parental obesity.* The New England Journal of Medicine, 1997.